PCT/GB 2003 / 0 0 0 9 8 4

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

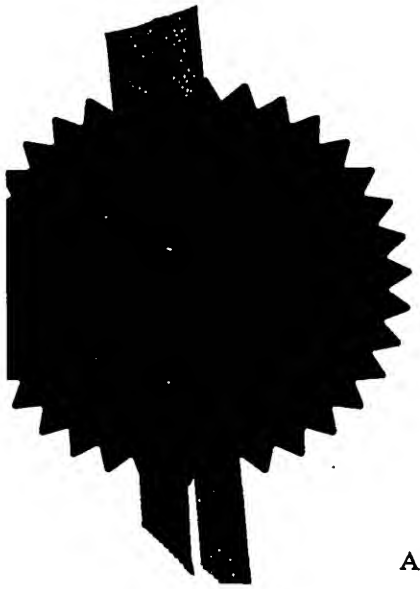
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed R. Mahoney

Dated 26 March 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

THE PATENT OFFICE
J
- 8 MAR 2002
LONDON
Fee: £0

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form.)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — HRW/VJR/44194 11MAR02 E702099-1 D01631 P01/7700 0.00-0205536.6

2. Patent application number *(The Patent Office will fill in this part)* — 0205536.6 — 8 MAR 2002

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

THE Babraham Institute
Babraham Hall
Babraham
Cambridge CB2 4AT
United Kingdom

Ae 11(7) 15/9/02

Patents ADP number *(if you know it)*
74145190001

If the applicant is a corporate body, give the country/state of incorporation — United Kingdom

4. Title of the invention — Tagging and Recovery of Elements Associated with Target Molecules

5. Full name, address and postcode in the United Kingdom to which all correspondence relating to this form and translation should be sent

Reddie & Grose
16 Theobalds Road
LONDON
WC1X 8PL

Patents ADP number *(if you know it)* — 91001 ✓

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application *(If you know it)*	Date of filing *(day/month/year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day/month/year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document.

Continuation sheets of this form	
Description	13
Claim(s)	3
Abstract	-
Drawing(s)	3

10. If you are also filing any of the following, state how many against each item.

Priority documents	-
Translations of priority documents	-
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	-
Request for preliminary examination and search *(Patents Form 9/77)*	1
Request for substantive examination *(Patents Form 10/77)*	-
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Reddie & Grose

Date 8 March 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

WAKERLEY, Helen Rachel
020-7242 0901

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Tagging and Recovery of Elements Associated with Target Molecules

The present invention relates to a new method for identifying elements associated with target molecules.

Many genes and gene clusters are controlled by known (or unknown) distant regulatory elements that are necessary for high-level expression. Identification of these regulatory elements is an expensive and time-consuming process. Previous attempts to identify such distant regulatory elements have used a number of different methods, but most directly by scanning large genomic regions for DNase I hypersensitivity sites, followed by functional analysis of those regions linked to reporter genes in transgenic mice. This method of identification will clearly take a very long time.

The beta-globin locus is the prototypical gene cluster regulated by distant regulatory elements; the search for the beta-globin regulatory elements took approximately 10 years. Experiments designed to locate the beta-globin gene regulatory elements began in the late 1970s. In the early 1980s data arose that suggested distant elements were involved. A thalassemia patient was studied whose genome contained an intact beta-globin gene but a large deletion upstream of the gene. This lead to the conclusion that a distant upstream element must be involved in the regulation of the gene (Kioussis et al., 1983). Indeed, transgenes containing the beta-globin gene alone achieve only very low levels of expression at best (Townes et al., 1985) In 1985 a series of DNase I hypersensitive sites were mapped 40-60 Kb upstream of the beta-globin gene (Tuan et al., 1985). In 1987 it was finally shown that this hypersensitive site region, collectively known as the locus control region (LCR), was

sufficient to induce high level, position independent, copy number dependent gene expression when linked to the beta-globin gene (Grosveld et al., 1987). Defects in human beta-globin gene expression, or hemoglobinopathies, are the most common genetic diseases worldwide. The ability to induce high-level expression of an artificially introduced beta-globin gene is therefore of significant therapeutic use. In addition, the ability to locate control regions of other genes is clearly desirable.

Chromatin conformation capture (3C; Decker et al 2002) has been used to determine the conformation of a yeast chromosome to try to determine the interaction of genes and control regions. However, many technical problems arise when trying to apply this method to higher eukaryotes, not least because the mammalian genome is approximately 200 times the size of a yeast genome. The 3C has several disadvantages: 3C does not enable recovery of in situ labelled molecules, nor does 3C give a very high degree of resolution. In addition, other disadvantages of the 3C technique result because this technique allows only an average conformation of a chromosome to be calculated; this means that if all the cells used in the technique are not homogeneous or the molecular conformation is dynamic, specific interactions may be overlooked. Further, the 3C technique does not provide a method for determining which proteins or other molecules are associated with the genome.

Fluorescence in situ hybridisation (FISH) is a previously known techniques which uses hapten-labelled nucleotide probes followed by anti-hapten antibodies conjugated to fluorophores to determine the site of an actively transcribed gene via the antibody's ability to specifically bind to the hapten. Covalent tag deposition has commonly been used to enhance the signals obtained using the above technique. Kits enabling performance of

covalent tag deposition to enhance signals are obtainable from NEN Dupont and are called TSA™ (Tyramide Signal Amplification™). However, this technique has not provided means for purifying molecular complexes from specific sites or in the immediate vicinity of specific sites in or on cells. Neither FISH nor TSA allow for detection (and thus identification) of, for example, the interaction of distant regulatory elements with an actively transcribed gene. There is no technique presently available to use for detecting(and thus identifying) the interaction of distant regulatory elements with an actively transcribed gene during the time of transcription.

Techniques are known which can be used for identification and analysis of proteins involved in protein complexes. ImmunoPrecipitation (IP) is most commonly used to 'pull down' proteins associated in a complex with a target protein(s). However no techniques exist to analyse, for instance, molecules or complexes which are only involved in "loose" functional interactions with another complex or which only function in the vicinity of another protein.

According to the present invention there is provided a method for identifying elements associated with target molecules comprising the steps of:
(a) providing a probe capable of binding specifically to a target molecule, the probe associated with an enzyme;
(b) adding a tag capable of being activated by the enzyme such that it can attach to elements in the vicinity of the enzyme; and
(c) isolating elements having the tag attached thereto.

The target molecules may include RNA molecules, DNA molecules, proteins or peptides, lipids, or other, artificial compounds.

When the target is RNA, the elements which may be associated with these target molecules and which may be identified (or whose mode of action can be understood) by using the technique of the present invention include: distant regulatory elements (i.e. DNA elements via their chromatin protein association) that are in proximity to the RNA of an actively transcribed gene; RNA binding proteins such as those involved in RNA processing or stabilization/regulation/etc; proteins and protein complexes which facilitate the interactions between regulatory elements and a gene; proteins and protein complexes involved in the activation of genes; proteins and protein complexes involved in the regulation of chromatin structure in and around active genes; and transcription factors.

When the target is DNA, the elements which may be associated with these target molecules and which may be identified (or whose mode of action can be understood) by using the technique of the present invention include: distant regulatory elements (i.e. DNA elements via their chromatin protein association) that are in proximity to the targeted DNA; other DNA elements in proximity to the targeted DNA, which may be for example, engaged in functional interactions with the target sequence (e.g. boundaries, insulators, structural or architectural interactions); analysis of higher order chromatin structure, for example the analysis of tertiary chromatin interactions (chromatin folding); mapping chromatin interactions in entire loci or whole genomes (with the aid of high throughput technology); protein/protein complexes involved in regulation of gene expression or the control of chromatin structure.

When the target is protein, the elements which may be associated with these target molecules and which may be identified (or whose mode of action can be understood) by

using the technique of the present invention include: DNA elements in proximity to a protein; RNA molecules in proximity to a protein; or other proteins/protein complexes bound to, or in the vicinity of a targeted protein (e.g. identifying other protein components of the LCR-beta-globin gene complex at different stages of development, or identifying the in-vivo ligands of a specific receptor- or vice versa).

When the target is lipid, the elements which may be associated with these target molecules and which may be identified (or whose mode of action can be understood) by using the technique of the present invention include: DNA elements in proximity to a lipid or artificial compound RNA molecules in proximity to a lipid or artificial compound; or proteins/protein complexes bound to, or in the vicinity of a targeted lipid or artificial compound

The probe usable in the present invention may be a DNA probe, an RNA probe or an antibody specific for a protein, lipid or other molecule.

The probes used can be associated with the enzyme through antibody/enzyme conjugates, or enzyme/target molecule fusion.

The method by which the enzyme may be targeted to a specific molecule may be varied depending on the molecule to be targeted. For example, using a labelled probe specific for a DNA molecule, using immuno-histochemistry, or using a fusion of a protein (or other molecule of interest) and the enzyme. Preferably antibody/enzyme conjugates may be used. In one preferred embodiment, when the target molecule is RNA, a hapten-labelled probe specific to the intron of an active gene can be added, followed by addition of a hapten-specific Fab

fragment/enzyme conjugate. One hapten which may be used is digoxygenin (DIG).

An enzyme which may be used in the present invention is Horse Radish Peroxidase. This enzyme can be used in combination with a tyramide molecule such as biotin-tyramide, dinitrophenol-tyramide or FITC-tyramide.

Another enzyme/TAG combination is ubiquitin-conjugating enzyme, with ubiquitin as a tag.

Protein kinase could also be used as the enzyme (there are several with varied specificities) with phosphate as a tag. In this example a kinase which is able to add a phosphate to a nucleosomal protein (if looking for chromatin tagging) or other protein of interest should be used. Antibodies against the specifically modified epitope of the particular amino acid residue receiving the phosphate could be used to target isolate the tagged elements.

DNA Adenine Methyltransferase (DAM) is another enzyme which could be used, with a methyl group as the tag. In a slight variation of the procedure, instead of using a tag to pull out the labelled material one could use a restriction enzyme that will cut only DNA which is specifically methylated by DAM. DAM adds a methyl group to the adenine in the sequence GATC. This methylated site can only be cut by the DNA restriction endonuclease DpnI. DAM is normally only found in bacteria such as E.coli so it could be used in eukaryotic cells without any interference from endogenous methyltransferases which only methylate other sequence combinations. With this method no affinity chromatography is required. We would simply purify the DNA from the DAM treated cells and cut with DpnI and then isolate small DNA fragments that are released from the mixture of genomic DNA. The small sites

released by DpnI digestion can then be labelled with radioisotopes, etc., and used for diagnostic hybridization to a microarray, for example (van Steensel et al 2001)

Other enzyme/tag combinations could be used: any enzyme which can activate a tag molecule to deposit onto another molecule, for example protein, DNA, RNA, lipid etc in a manner such that the tagged product can then be isolated by whatever means (eg. affinity chromatography or immunoprecipitation) can be used in this technique.

Before separation, the molecules which have been tagged can be disrupted into smaller fragments using, for example, sonication, enzymatic cleaving, shearing with a French Press or small bore syringe, or another method which achieves such a result.

Analysis of the DNA obtained using the above method can be used to identify any regulatory elements which were in proximity to the active gene, because these elements become labelled with the tag, due to their proximity to the site HRP activity. The DNA can then be analysed by a number of quantitative techniques, for example Quantitative PCR (for example Real-Time PCR (Wittwer et al., 1997)) or semi-quantitative PCR, slot blot or microarray (Granjeaud et al., 1999), among others. This analysis allows scanning, high-throughput, high resolution analysis of any gene locus for hundreds or thousands of kilobases in either direction.

An embodiment of the present invention will now be described in more detail, by way of example, with reference to the drawings, in which:

Figure 1 is a schematic diagram showing a method of the present invention

Figure 2 is a schematic diagram showing the mouse beta-globin locus and locus control region (LCR) and two models of LCR action; and

Figure 3 is a schematic diagram showing the hypothesised interaction of the mouse beta-globin locus and locus control region (LCR), as a result of performing the present invention.

Many genes and gene clusters are thought to be regulated by distant regulatory elements, which may be located tens to hundreds of kilobases away. The best characterised example of a distant element regulating a cluster of genes is the beta-globin locus control region (LCR), shown in Figure 1. The LCR **7** consists of a series of Dnase I hypersensitive sites (HS) (**1 to 6**). At the core of each HS is a 200-300 bp region which is packed with transcription factor binding sites. The LCR is absolutely required for high level transcriptional activation of all the beta-globin genes. Two models have been proposed to explain the action of the LCR, although no direct proof exists for either mode of action, these are shown in Figure 2. The first model **8** proposes that the LCR works at a distance. The LCR creates a large region of open chromatin surrounding the genes and recruits and sends factors necessary for gene activity along the chromatin. The second model **9** proposes that the LCR physically contacts the gene(s) through long range chromatin interactions, essentially looping out the intervening sequences and activating transcription directly.

To determine if an actively transcribed beta-globin gene is in direct physical contact with the distant (40Kb) LCR in vivo, the following technique was used (see Figure 2). Firstly, fetal liver **10**, the main site of erythropoiesis in the developing foetus, is taken and disrupted, and the

cells are spread in a monolayer on a slide **11**, prior to cross-linking with formaldehyde. In situ hybridization is performed using a digoxygenin (DIG)-labelled oligonucleotide probe **12**, specific for the intron of the mouse beta-major globin gene. The enzyme Horse Radish Peroxidase (HRP) is then targeted to an RNA molecule using an anti-DIG antibody conjugated to Horse Radish Peroxidase (HRP) **13**, thus pinpointing HRP enzyme activity to the site of the actively transcribed gene.

Next, biotin-tyramide **14** is added as a molecular tag; it is activated by the HRP to cause it to covalently attach to electron dense amino-acids in the immediate vicinity. After the tag is covalently attached **15**, the cells are sonicated to give small, soluble chromatin fragments **16** having an average DNA size of 400bp. The biotinylated chromatin is then purified using streptavidin agarose affinity chromatography **17**, cross-links are reversed and the DNA is purified. Multiple amplicons across the locus can then be analysed **18** using quantitative or semi-quantitative PCR and/or slot blotting.

By using the above technique on the mouse beta-globin gene locus, it was found that high-level expression of the beta-globin genes is totally dependent on an extensively characterised, distal, regulatory element known as the LCR. The LCR and active beta-major gene are found to be in significant proximity in the mouse beta-globin locus in vivo; HS2 **2** appears to be in intimate contact with the beta-major gene, and the two active adult genes also appear to be in close proximity (Figure 3).

There are many applications for the technique of the present invention, which can be performed in vivo, ex vivo, or in vitro.

One example of such a use is in transgenic animal technology: transgenic animals are presently being used by a number of laboratory around the world as bioreactors to produce large amounts of proteins of interest. The most commonly used method is to express the protein of interest in milk under control of a highly expressed milk protein gene promoter. Most transgenic animals created with such a construct would not express the protein or express it at very low levels making them unusable. Some transgenic animals may, by virtue of position effects at the site of integration of the construct, express larger amounts of the protein of interest. The addition of milk protein gene LCR-like sequences to the expression construct would increase the number of transgenic animals which express the gene to 100% and increase the average level of expression in every animal. This would significantly decrease the cost of production and greatly increase the yield.

When RNA is the target molecule, the method of the present invention labels only the cells in the population that are actively transcribing the gene of interest. The advantage of this is specifically interacting sequences are highly enriched upon affinity chromatography, whether the population is heterogeneous or the interaction is dynamic (Wijgerde et al., 1995). Another advantage of the present invention when RNA is the target molecule is this technique can detect (and thus identify) the interaction of distant regulatory elements with an actively transcribed gene during the time of transcription. There is no other technique we know of which can be used for this purpose. This technique can specifically label and recover proteins at the site of transcription in a dynamic or heterogeneous population of cells and identify specific interactions.

Another advantage of the present invention which results whatever the target molecule is, is the possibility of labelling and recovering complexes in the vicinity of a target complex (as opposed to molecules which are in direct interaction). The resultant enriched proteins could be analysed by a number of protein chemistry techniques such as Western blotting, Mass Spectroscopy, fractionation, purification, polyacrylamide gel electrophoresis, etc.

The present invention provides a relatively easy and rapid method which can detect interactions between an actively transcribed gene and distant regulatory element(s). The technique can also be used to identify any sequence element involved in an interaction with any other target sequence in vivo by virtue of their proximity.

The present invention provides a new way to identify the regulatory elements involved in the activation of genes in a rapid and relatively inexpensive way. It has also been used to address the question of how LCRs or enhancer elements function and in fact has provided the first direct evidence that the LCR functions by physically interacting with an actively transcribed gene in the beta-globin locus.

REFERENCES

Bobrow M., Harris, T., Shaughnessy,K. and Litt, G. Catalyzed reporter deposition, a novel method of signal amplification - application to immunoassays. Journal of Immunological Methods 125: 279-285 1989

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science 295, 1306-11.

Granjeaud, S., Bertucci, F., and Jordan, B. R. (1999). Expression profiling: DNA arrays in many guises. Bioessays 21, 781-90.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51, 975-85.

Kioussis, D., Vanin, E., deLange, T., Flavell, R. A., and Grosveld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta- thalassaemia. Nature 306, 662-6.

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L., and Palmiter, R. D. (1985). Erythroid-specific expression of human beta-globin genes in transgenic mice. Embo J 4, 1715-23.

Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985). The "beta-like-globin" gene domain in human erythroid cells. Proc Natl Acad Sci U S A 82, 6384-8.

van Steensel, B., Delrow, J. and Henikoff, S. Chromatin profiling using targeted DNA adenine methyltransferase. Nature Genetics Volume 27 March 2001

Wijgerde, M., Grosveld, F., and Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. Nature 377, 209-13.

Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. (1997). Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22, 130-1, 134-8.

Claims

1. A method for identifying elements associated with target molecules comprising the steps of:

(a) providing a probe capable of binding specifically to a target molecule, the probe associated with an enzyme;

(b) adding a tag capable of being activated by the enzyme such that it can attach to elements in the vicinity of the enzyme; and

(c) isolating elements having the tag attached thereto.

2. A method as claimed in claim 1 in which the target molecule is selected from the group consisting of RNA molecules, DNA molecules, proteins or peptides, lipids, or other, artificial compounds.

3. A method as claimed in claim 1 or 2 in which the elements which may be associated with the target molecules include distant regulatory elements, RNA, DNA, proteins and protein complexes, transcription factors, or in-vivo ligands of a specific receptor.

4. A method as claimed in any preceding claim in which the probe is selected from the group consisting of DNA probe, an RNA probe or an antibody specific for a protein, lipid or other molecule.

5. A method according to claim 4 in which the probe is associated with the enzyme through an antibody/enzyme conjugate, or enzyme/target molecule fusion.

6. The method according to any preceding claim in which the enzyme is targeted to RNA using a hapten-labelled probe specific to the RNA of an intron of an active gene, and then a hapten-specific Fab fragment/enzyme conjugate is added.

7. The method according to any preceding claim in which the enzyme is Horse Radish Peroxidase and the tag is biotin-tyramide.

8. The method according to any preceding claim in which elements are isolated using affinity chromatography or ImmunoPrecipitation.

9. A method for identifying elements of chromatin associated with transcribing RNA comprising the steps of:

(a) providing a hapten-labelled probe capable of binding specifically to RNA of a gene,

(b) providing an antibody conjugated with the enzyme horse-radish peroxidase, the antibody specific for the hapten;

(c) adding biotin-tyramide such that it can attach to elements in the vicinity of the enzyme;

(d) disrupting the chromatin

(e) isolating elements of chromatin having biotin attached thereto using affinity chromatography and purifying the elements.

10. The method of claim 9 in which the chromatin is disrupted using sonication, enzymatic cleaving, or shearing with a French Press or small bore syringe.

11. The method according to any of claims 7 to 9 in which the hapten is digoxygenin.

12. Elements isolated by the method of any preceding claim.

13. Analysis of DNA obtained using the method according to any preceding using Quantitative Real-Time PCR, slot blot or microarray.

14. A method for identifying DNA associated with target molecules comprising the steps of:

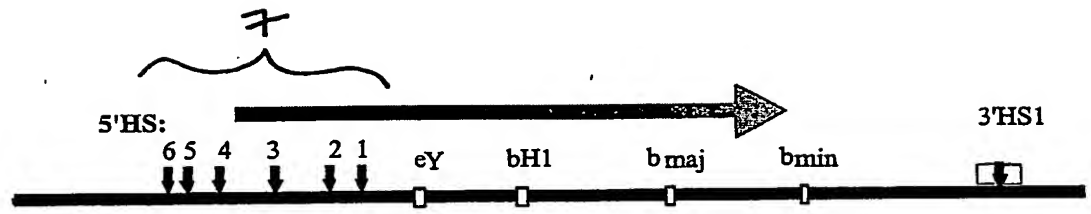
(a) providing a probe capable of binding specifically to a target molecule, the probe associated with an DNA Adenine Methyltransferase;

(b) adding a restriction enzyme that will cut only DNA specifically methylated by DAM;

(c) isolating DNA cut by the restriction enzyme

(d) identifying the isolated DNA.

B



C

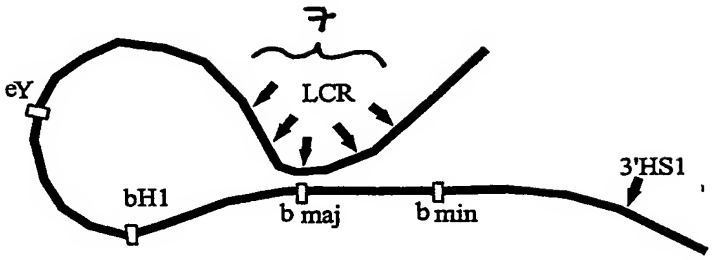


FIGURE 1

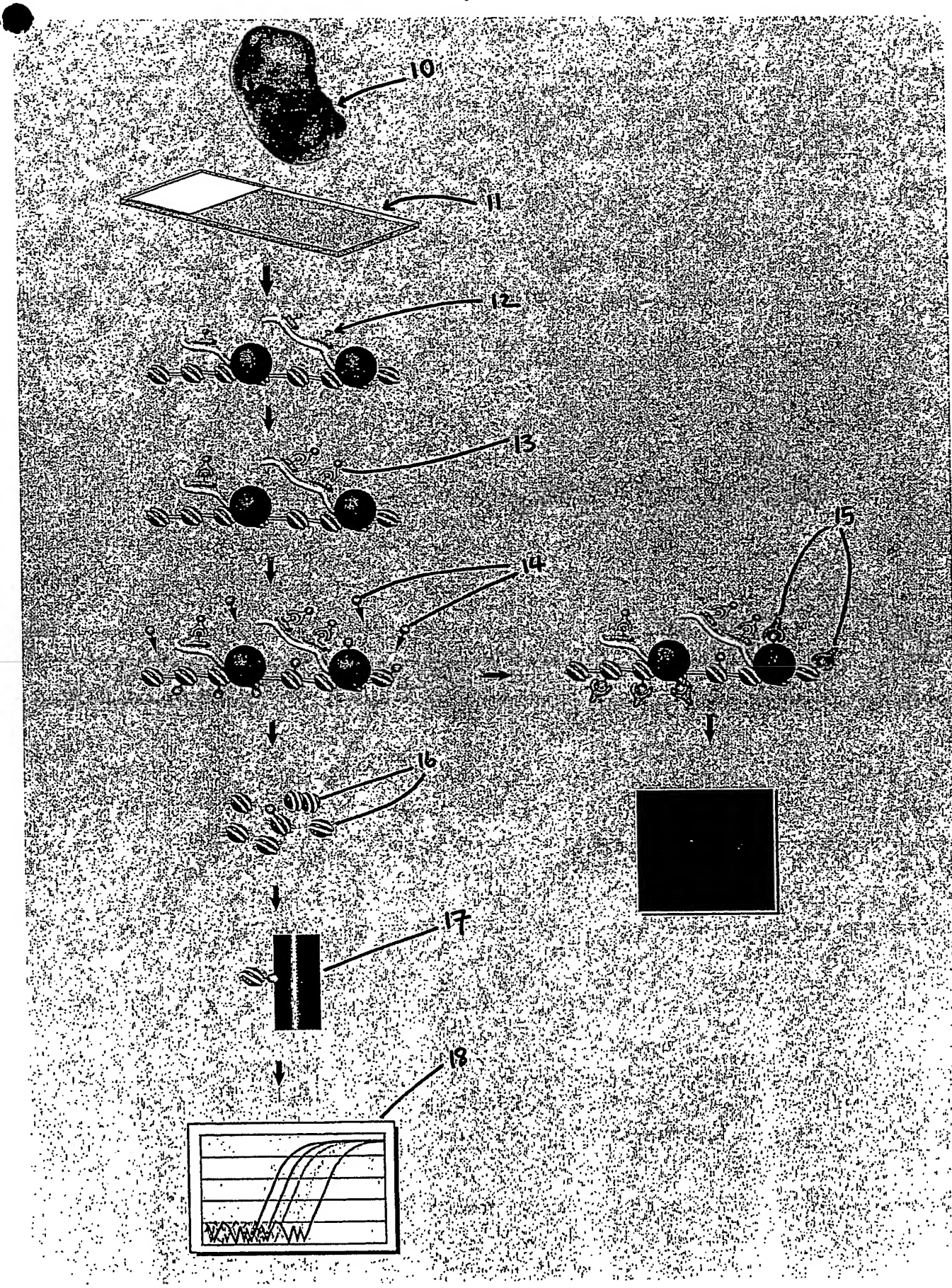


FIGURE 2

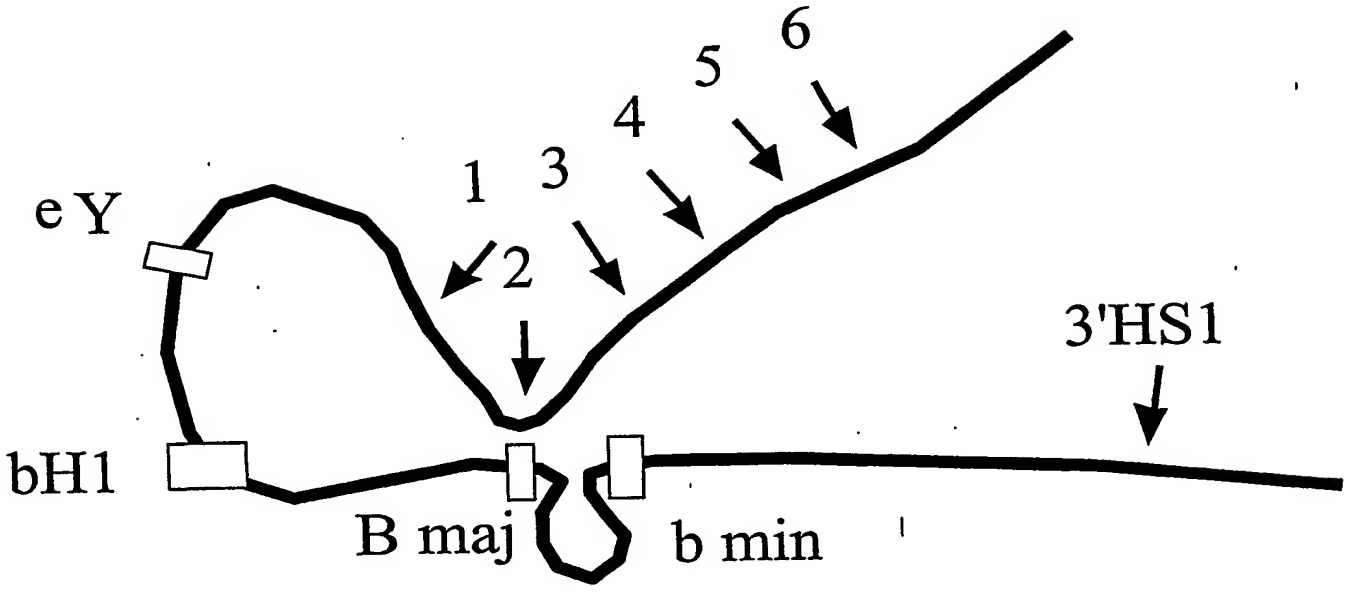


FIGURE 3